International Telegraph and Telephone Consultative Commitee (CCITT)

Question 4/XV Specialists group on coding for visual telephony

Date: 20 september 1985

Source: SWEDEN

Title: **A COMMENT ON ENTROPY ESTIMATION**

The aim of estimating entropy is to get a measure of lowest possible bit rate after optimal variable length coding. A recent theoretical study [1] shows that entropy estimation is a delicate task. On the other hand, bit rate estimation is more tractable. The derivation of the below main results can be found in [1].

Consider a random variable with N possible outcomes, having probabilites

$$P = (p_1, p_2, \ldots, p_N)$$

and entropy

$$H = \sum_{i=1}^{N} p_i \log \frac{1}{p_i}$$

Assume further we observe n outcomes of the r.v., each letter occurring k_i times. The mean square, unbiased esimates of P and H are then

$$\hat{p}_i = \frac{k_i + 1}{n + N}$$

$$\hat{H} = \frac{1}{\ln 2} \sum_{i=1}^{N} \hat{p}_i \sum_{j=k_i+2}^{n+N} \frac{1}{j}$$

Using arithmetic coding, the bit rate estimate after variable length coding becomes

$$\hat{R} = \sum_{i=1}^{N} \hat{p}_i \log \frac{1}{\hat{p}_i}$$

[1] H.Brusewitz "Estimation of Probabilities and Entropy" Report no TRITA-TTT-8406, dec 1984